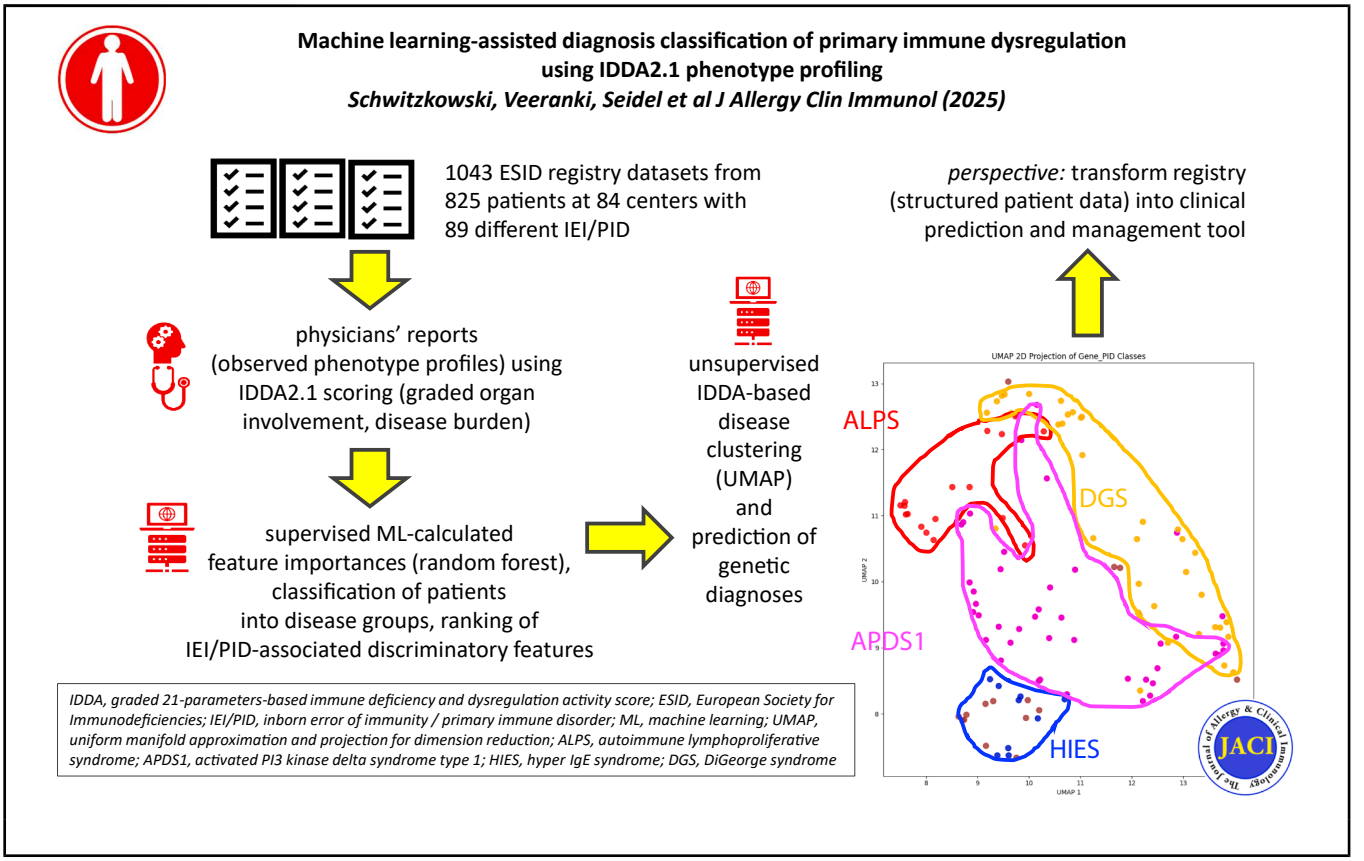
Machine learning–assisted diagnosis classification of primary immune dysregulation using IDDA2.1 phenotype profiling

Malte Schwitzkowski, BSc, Sai Pavan Kumar Veeranki, DDI, Benedikt N. Seidel, BSc, Gerhard Kindle, MD, Dipllnf, Stephan Rusch, Dipllnf, the European Society for Immunodeficiencies Registry Working Party, et al

GRAPHICAL ABSTRACT



Capsule summary: Phenotype profiles of 825 patients with inborn errors of immunity/primary immunodeficiency (IEI/PID) with a focus on immunodysregulation were collected in a European Society for Immunodeficiencies registry study for machine learning–assisted disease classification. Random forest analyses ranked feature importances, and uniform manifold approximation and projection allowed clustering of patients reflecting their IEI/PID.

J Allergy Clin Immunol 2026;157:470-85.

https://doi.org/10.1016/j.jaci.2025.10.022

Machine learning–assisted diagnosis classification of primary immune dysregulation using IDDA2.1 phenotype profiling

Malte Schwitzkowski, BSc,[a] Sai Pavan Kumar Veeranki, DDI,[b] Benedikt N. Seidel, BSc,[a,c] Gerhard Kindle, MD, Dipllnf,[d,e,f] Stephan Rusch, Dipllnf,[d,e] the European Society for Immunodeficiencies Registry Working Party,* Diether Kramer, PhD,[b] and Markus G. Seidel, MD[a] *Graz, Austria; and Munich and Freiburg, Germany*

Background: Immune dysregulation, including autoimmunity, autoinflammation, allergy, and malignancy predisposition, adds significant disease burden in primary immune disorders (PID) and inborn errors of immunity (IEIs).
Objective: We evaluated whether the 5-graded immune deficiency and dysregulation activity (IDDA2.1) score, encompassing 21 organ involvement and disease burden parameters, supports diagnosis across a wide spectrum of IEIs.
Methods: From April 2022 to November 2024, collaborators from 84 centers collected 1,043 IDDA score datasets from 825 patients across 89 IEIs (17 disorders with ≥10 patients each; range, 1-196 per IEI), including 177 scores from 141 treated patients. Supervised machine learning models (*k*-nearest neighbors, support vector machine, logistic regression, random forest) classified patients into disease groups and ranked corresponding predictive features, while unsupervised uniform manifold approximation and projection (UMAP) visualized disease-specific clustering.
Results: Feature analysis reflected clinicians' recognition of IEI patterns and confirmed internal IDDA score consistency. Phenotype profiles in treated patients remained informative, inversely reflecting anticipated treatment-dependent phenotype amelioration. UMAP effectively distinguished IEIs by IDDA2.1 profiles. Genetic disorder prediction achieved 73% overall accuracy, 70% for the correct monogenic IEI, and 93% within the top 3 predictions; classification reached 43% for IEI–International Union of Immunological Society categories and 59% for 12 "cardinal" IEIs (25 genes).
Conclusions: Random forest feature importance analysis can inform targeted clinical screening for key disease manifestations. The top 3 prediction approach demonstrates diagnostic potential, but improved accuracy will require larger, globally shared datasets. Small sample sizes for rare diseases highlight the necessity of broader collaboration to enhance AI-assisted clinical decision-making in the future. (J Allergy Clin Immunol 2026;157:470-85.)

Key words: *Inborn error of immunity (IEI), primary immunodeficiency (PID), primary immune regulatory disorder (PIRD), phenotype-driven disease classification, interoperable patient data, immune deficiency and dysregulation activity (IDDA) score, artificial intelligence (AI), unsupervised and supervised machine learning (ML), primary immune disorder (PID)*

Inborn errors of immunity (IEIs) may present with infection, autoimmunity, autoinflammation, allergy, and malignancy, and any of these features may manifest initially or later during their clinical course.[1-7] The degree and type of features of immune dysregulation like autoimmune cytopenia, lymphoproliferation, enteropathy, among others, depend on the imbalance of effector and regulatory immune functions caused by the underlying defect, and they are present in at least a quarter of primary (PID) or secondary immune disorders.[8,9] Nevertheless, even in "diseases of immune dysregulation," listed in table 4 of the International Union of Immunological Societies' classification of IEIs,[5] including monogenic primary immune regulatory disorders, there is rarely a clear genotype–phenotype correlation that reliably defines the type and severity of features of immune dysregulation to be expected in a patient.[10-12] The phenomenon of highly varying extent, and timing of the onset of immune dysregulation in IEI/PID may be due to incomplete penetrance,[13] untimely infectious triggers, or other genetic or epigenetic cofactors. It may lead to mortality or substantial morbidity in one individual while seemingly unaffected, asymptomatic siblings or parents carrying the same pathogenic variant are present. It makes it difficult for clinicians to predict outcomes or to decide on adequate diagnostic tests and the intensity and timing of (definitive) treatments. Furthermore, the evaluation of responses to treatment measures in immune dysregulation may be complicated by varying organ involvement and their potentially different sensitivity to treatment.

From [a]the Styrian Children's Cancer Research Unit for Cancer and Inborn Errors of the Blood and Immunity in Children, Division of Pediatric Hematology and Oncology, Department of Pediatric and Adolescent Medicine, Medical University of Graz, Graz; [b]the Styrian Hospital Association (KAGes), Directorate for Technology and IT, IT Management Team, Graz; [c]the Autonomous Aerial Systems, School of Engineering and Design, Technical University of Munich, Munich; and [d]the Institute for Immunodeficiency, Center for Chronic Immunodeficiency, [e]the Centre for Biobanking FREEZE, Medical Center, and [f]the Treuhandstelle, Trusted Third Party, Faculty of Medicine, Medical Center, University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg.

*Members of the European Society for Immunodeficiencies Registry Working Party are listed in the Acknowledgments.

The first 3 authors contributed equally to this article, and all should be considered first author.

This article is part of a special issue entitled: Artificial Intelligence (AI) in Allergy/Immunology published in the *Journal of Allergy and Clinical Immunology.*

Received for publication April 17, 2025; revised October 6, 2025; accepted for publication October 10, 2025.

Available online November 5, 2025.

Corresponding authors: Markus G. Seidel, MD, Division of Pediatric Hematology and Oncology, Medical University of Graz, Auenbruggerplatz 38, 8036 Graz, Austria. E-mail: markus.seidel@medunigraz.at. Or: Sai Pavan Kumar Veeranki, DDI, Styrian Hospital Association (KAGes), Directorate for Technology and IT, Billrothgasse 18a, 8010 Graz, Austria. E-mail: Sai.Veeranki@kages.at.

The CrossMark symbol notifies online readers when updates have been made to the article such as errata or minor corrections

0091-6749

https://doi.org/10.1016/j.jaci.2025.10.022

Abbreviations used

AI: Artificial intelligence
ALPS: Autoimmune lymphoproliferative syndrome type 1
APDS: Activated PI3-kinase delta syndrome
AT: Ataxia-telangiectasia
BTK: Bruton tyrosine kinase
CTLA-4: Cytotoxic T lymphocyte–associated protein 4
CVID: Common variable immunodeficiency
DGS: DiGeorge syndrome (aka 22q11.2 deletion syndrome)
ESID: European Society for Immunodeficiencies
HIES: Hyper-IgE syndrome
ICU: Intensive care unit
IDDA2.1: Immune deficiency and dysregulation activity v2.1
IEI: Inborn error of immunity
IGRT: Immunoglobulin replacement therapy
IUIS: International Union of Immunological Societies
LIME: Local interpretable model-agnostic explanations
LRBA: LPS-responsive and beige-like anchor protein
MDI: Mean decrease in impurity
ML: Machine learning
NBS: Nijmegen breakage syndrome
NFKB1/2: Nuclear factor–kappa beta subunit 1/2
PID: Primary immune disorder
RF: Random forest
UMAP: Uniform manifold approximation and projection

To address the challenge of a qualitative and quantitative assessment of organ involvement, disease burden, and responses to therapy in patients with deficiency of lipopolysaccharide-responsive beige-like anchor protein (LRBA deficiency), we developed the immune deficiency and dysregulation activity (IDDA) score.[14] The original intention of using this score was to document, in a simple fashion, by grading, from 0 to 4, the severity of noninfectious involvement of every organ then known to be possibly affected in LRBA deficiency, thereby taking a snapshot of the degree of immune dysregulation. Additionally, other quantitative factors, such as performance scales and intensive care unit (ICU) or general hospital stays (in percentages), and the need for supportive care, were documented in a semiquantitative way. These were used together with organ grading to generate a numerical output that could be applied to compare a patient's disease burden cross-sectionally with others as well as to monitor the same patient longitudinally over various phases of the disease and treatment.[14] Recently it has been shown that the immune deficiency and dysregulation activity v2.1 (IDDA2.1) score was markedly reduced after successful hematopoietic stem cell transplantation in patients with cytotoxic T lymphocyte–associated protein 4 (CTLA-4) haploinsufficiency. Moreover, low pretransplantation IDDA scores correlated significantly with higher overall and disease-free survival probabilities.[15] In addition to the calculated numerical (total) IDDA score itself that represents a one-dimensional quantitative measure, documenting the multiparametric score in individual patients repeatedly over time allowed us to compare treatment responses of each organ manifestation individually (eg, before-and-after grading per organ), creating an immune dysregulation–based phenotype profile of the disease state.[14] Furthermore, to be able to apply the IDDA score to other IEI/PIDs with immune dysregulation more generally, an amended version IDDA2.1 that included hemophagocytosis as additional possible phenotypic feature, a slightly clarified wording of definitions, and an improved calculation formula, putting less weight on performance scales than before, was designed.[16] The application of the IDDA2.1 version to evaluate the organ involvement of a variety of IEI/PID with immune dysregulation allowed phenotype profiling, then termed *kaleidoscope scoring* because of the disease-characteristic patterns of spider web plots generated from the data.[16] However, the limitation of that approach was that at that time, only cumulative qualitative (aggregate) data of organ involvement probabilities were available from published patient series and reviews of 18 selected diseases retrospectively, and the real-life occurrence of immune dysregulation in IEI/PID and per-patient graded (quantitative individual) phenotype profiles other than from the original cohorts of patients with LRBA deficiency or CTLA-4 haploinsufficiency were lacking.

With the recent developments in machine learning (ML), a foreseen potential application of the IDDA phenotype profile is to feed it into pattern recognition algorithms in order to learn and predict the diagnosis of a hitherto undiagnosed patient. Consequently, such an early classification based on similarities of IDDA phenotype profiles to those of patients with known monogenic IEI/PID might allow clinicians to plan diagnostic screening tests and take precautions for expectable, disease-specific involvement of risk organs (eg, lung in LRBA deficiency[13]) as well as early evaluation of indications for phenotype-targeted treatment.

In line with the rapidly growing use of AI applications in medicine,[17] various approaches have been made in the recent years to accelerate diagnosis making in IEI/PID from directly interpreting electronic health records by using AI applications like natural language processing and large language models.[18-21] Recently a ML algorithm that learned phenotypic patterns from electronic health records of patients with common variable immunodeficiency (CVID) to rank the likelihood of undiagnosed patients for their probability to have CVID was shown to greatly potentially reduce the diagnostic delay in the majority of patients.[22]

We hypothesized that data derived from IEI/PID patient registries such as the European Society for Immunodeficiencies (ESID) registry were prestructured and contain expert-specified qualitative, and in part quantitative, phenotypic details. These data, together with IDDA2.1 data, represented a perfect source to train an AI model to classify and predict IEI/PID diagnoses and their manifestations. Moreover, using the IDDA score in a large number of patients in the context with AI applications to analyze them would enable identifying disease-specific profiles of patients with IEI/PID and immune dysregulation. To collect unbiased data from a wide variety of diagnosed IEI/PID with immune dysregulation and validate the IDDA2.1 phenotype profile and score on a maximum number of patients, we launched an ESID registry study in April 2022. We aimed to test the utility of real-word IDDA2.1 parameter profiles of patients with any IEI/PID with immune dysregulation using datasets of partly genetically diagnosed patients for ML-based disease classification. Additionally, the association of IDDA2.1 score with the phenotype as well as the contribution of other documented features were evaluated. Furthermore, dimensionality reduction was performed to project the high-dimensional data into a low-dimensional space, thereby positioning similar patients in close proximity, enabling the identification of intrinsic patterns and clusters.

METHODS

Patients

The patient data were acquired by retrospective chart review from prospective monitoring via the optional ESID level 2 registry "IDDA Score Module" (see Fig E1 in this article's Online Repository available at www.jacionline.org) from 84 participating centers across Europe and neighboring countries between April 2022 and November 2024. The end date chosen for data inclusion was November 9, 2024. Kindle et al[7] provide further details of the ESID registry. Any patient with features of immune dysregulation could be included, without restriction regarding IEI diagnosis or International Union of Immunological Society (IUIS) category, except if prior definitive treatment such as hematopoietic stem cell transplantation or gene therapy had been provided. There were no further explanations for IDDA scoring other than that shown in the registry module (Fig E1). Simultaneously, any immune-modifying treatment provided during the last 3 months was captured qualitatively. A yearly follow-up was expected for patient data entries in the ESID registry, but an unlimited number of additional entries (ie, more frequent follow-up) was possible. The study protocol, the patient informed consent template, and the center data transfer agreement with amendments were approved by the institutional review board and the data protection officer at the Medical University of Freiburg, Germany (Albert-Ludwigs-University). A substantial amendment was approved by the Medical University of Graz, Austria, in 2024 (24-334 ex 11/12, IRB00002556).

Data and preprocessing

The original dataset included the 21 parameters of the IDDA2.1 kaleidoscope score and the calculated score (total) including hospitalization metadata and performance scales. Each data entry was associated with patient age and sex. Disease labels encompassed genetic defects by the abbreviation used in the ESID registry, the IEI/PID diagnosis names, and IUIS classification main categories. For some analyses, either the genetic diagnosis or, if unavailable, the clinical diagnosis was used (labeled "gene_pid"), where an existing genetic diagnosis was prioritized over a clinical diagnosis—for example, if nuclear factor–kappa beta subunit 1 (NFKB1) deficiency was prioritized over CVID. An arbitrary label, *cardinal IEI,* was introduced in an attempt to distinguish IEIs with supposedly highly characteristic patterns of immune dysregulation (12 IEIs or subgroups of biologically similar IEIs with 25 different underlying genetic defects).

Multiple hierarchies of diseases were used as labels, such as IUIS tables,[5] names of IEI/PID, and genetic defects, as well as cardinal genes and the general appearance of a known genetic defect. Thus, the final dataset contained 24 features and 6 label sets. Three subcohorts were chosen to be analyzed: (1) the full dataset, (2) all patients at their initial consultations, and (3) initial consultations without immune-modifying therapy. The first group contained the first 14 parameters of the IDDA2.1 score. The second group contained more general information about immunoglobulin replacement therapy (IGRT), infections, organ dysfunctions, and nutrition status. The third group contained time spent in the hospital (general ward or intensive care unit) and the Karnofsky or Lansky performance scale scores. Although the IDDA score parameters for performance scale, hospitalization, or ICU stays (percentage of days in the latest observation period; eg, number of days out of the last 100 days) are helpful in the assessment of an individual's overall disease activity and burden and for its longitudinal monitoring, these 3 parameters were found to be rather patient than disease specific, especially when looking at small patient cohorts like in the present study. Thus, this group of 3 parameters was omitted from some of the subanalyses, yielding an 18-parameter set. Furthermore, the IDDA score itself was introduced as a feature group and demographic information such as sex and age as the last group. Various combinations of the feature groups were analyzed (see Table E1 in the Online Repository available at www.jacionline.org).

Each dataset was split into training and testing datasets in a 3:1 ratio, which was based on considerations of classes with small sample size to ensure that the performance of the classifier models could be validated with at least one sample for each class. We note that the datasets had been derived from 84 independent participating centers, and test datasets were never used for training. Two approaches were considered for handling small class sizes—that is, classes with fewer samples than necessary, both (1) excluding extremely small classes and (2) merging them into a single group.

Classification

The classification experiment, aiming to annotate the correct class (IEI diagnosis, IEI category, or genetic defect) to a dataset based on IDDA score phenotype parameters/profiles, was designed as multiclass classification task,[23-25] using classical ML algorithms including elastic net logistic regression,[26] support vector machine,[27] *k*-nearest neighbor,[28] and random forest (RF).[29] The main goal was to compare different classifiers; the best performance models are presented in the Results. Binary classification selects one of two classes, multiclass selects one of many, and multilabel assigns multiple classes to one sample.

Hyperparameter optimization was performed by using a grid search with 10-fold cross-validation. Given the class imbalance (substantially varying cohort sizes of groups with rare IEI diagnoses or genetic defects, complicating the weighting of results), balanced accuracy was considered for optimization. For *k*-nearest neighbor analysis, we optimized the number of neighbors, weighting method, and distance metric. The support vector machine was tuned for the regularization parameter *(C),* kernel type, degree, and gamma. Logistic regression was adjusted for penalty type, L1 ratio, solver, and maximum iterations. For RF, we optimized the number of trees, maximum depth, and minimum samples per split, along with the number of features considered at each split.[30] Hyperparameter optimization tunes model parameters iteratively to achieve the best performance. Cross-validation is a bootstrapping technique to enable the iterations for hyperparameter optimization; the actual training data are split into several portions, and each portion acts as a test set while the rest of the portions act as training set to determine the performance measure (eg, accuracy).

To identify the most influential features (single IDDA score parameters) associated with each disease class, we used feature importance analysis by using the RF algorithm in a one-versus-all multiclass classification setup. The most relevant features across all diseases were determined by comparing the mean decrease in impurity (MDI), which measures how much each feature reduced uncertainty, on average, when used to split data in a decision tree (the higher, the more relevant) in RF's "Feature Importance." This approach involves training a separate binary classifier for each class, treating the target class as positive and all others as

negative. By doing so, we can isolate the contribution of each feature to the prediction of a specific label (IEI diagnosis, category, or gene). RF,[29,31,32] with its ensemble of decision trees, naturally provides a measure of feature importance that is based on an impurity function that reduces the ambiguity in selecting the class. This enables a robust estimation of which features are most critical for distinguishing each class from the rest, offering valuable insights into class-specific patterns and aiding in the interpretability of the model's decisions.[32,33] In addition, patterns of features with high MDI were compared visually. To interpret individual predictions, we used the LIME (local interpretable model-agnostic explanations) technique.[34] Feature importance is a way of measuring which input factors (or features—in our case, single parameters of the composite IDDA score) have the strongest influence on the model's predictions.

We further explored how IDDA score parameter profiles related across patient groups using a technique called uniform manifold approximation and projection (UMAP).[35] UMAP is a tool that helps reduce complex data into a simple 2- or 3-dimensional visual format that humans can easily interpret. It is particularly useful when working with medical data that includes many different variables because it allows us to visualize how individual patient profiles relate to each other. A UMAP web application was deployed to enable health care professionals to explore clustering and separation patterns of various genetic defects interactively (see the web application available at https://esid.org/html-pages/IDDA_AI_SupplFigure3_UMAP_gene_pid_setup_2.html).

The evaluation of the model performance was performed by analyzing accuracy, precision, recall, F1 score, balanced accuracy, and area under the receiver operating characteristics curve.[36] Although the experiments were designed as multiclass classification in order to account for discrepancies in predicted probabilities that could lead to misclassification, we additionally evaluated performance on the basis of whether the true label was among the classes with top 3 predicted probabilities.[31]

Code and graphics

Python[37,38] and R[39] programing languages were used to perform several tasks for this study. Package Pandas, SciPy,[38] and 'numpy'[40] in Python and 'dplyr'[41] in R were used for data preprocessing. Scikit-learn[42] in Python and 'caret,' 'randomForest,'[43] and 'pROC'[44] in R were used for applying ML methods and calculate performance measures. Matplotlib[45] in Python and 'ggplot2'[46] in R were used to generate graphics. The R packages 'umap'[47] and 'plotly'[48] were used to generate interactive HTML to play with the UMAP map.

RESULTS

Dataset and patient cohort

We received 1,062 IDDA score data entries for 843 patients. Patients without diagnostic information (18 datasets) and one entry with corrupted treatment data were excluded, resulting in 1,043 valid entries for 825 individual patients, including 138 patients with 219 follow-up entries (Table I). Of those, 725 patients were documented to be untreated at 913 visits. There was a mild male predominance (457 male vs 368 female participants), and the median (range) age was 22.7 (0.2-84) years. The diagnosis distribution of patients with 89 different IEIs recorded is listed in Table I. While the large majority of patients had CVID without a genetic diagnosis, we received many entries of patients with specific monogenic IEIs, such as activated PI3-kinase delta syndrome (APDS) types 1 and type 2, DiGeorge syndrome (DGS, aka 22q11.2 deletion syndrome), ataxia-telangiectasia (AT), Bruton tyrosine kinase (BTK) deficiency, CTLA-4 haploinsufficiency, or autoimmune lymphoproliferative syndrome type 1 (ALPS; Table I), among others.

Clinical (human) phenotype profiling

We first calculated the means of available IDDA scores per parameter per disease and plotted 18 of 21 parameters separately for untreated subjects and for patients receiving immune-modifying therapy. We omitted the parameters for performance scales and ICU/hospitalization frequencies from these and clustering analyses because we noted that in the small patient subcohorts available (mostly <10-30 individuals), single patients with prolonged hospital stays or poor performance scales substantially skewed the dataset. A selection of IDDA2.1 kaleidoscope phenotype profiles recorded in this study is shown in Fig 1. Resembling physicians' clinical pattern recognition, diseases with biological similarity showed highly similar phenotype profiles (eg, CTLA-4 haploinsufficiency and LRBA deficiency; Fig 1, *A* and *B*), whereas they differed markedly from patients with ALPS, DGS, hyper-IgE syndromes (HIESs), or APDS (Fig 1, *C-F*). We separately analyzed patients receiving immune-modifying therapy (eg, mycophenolate mofetil, sirolimus, leniolisib), expecting ameliorated or unspecific phenotype profiles due to treatment. Although only a small number of datasets was available, patients receiving therapy were documented as having more pronounced IDDA scores (increased values were higher than in untreated patients, and some of the low values were lower; Fig 1, *right*), suggesting possible biases.

ML-based phenotype profiling

To assess the relative relevance of the 21 individual IDDA parameters and the calculated IDDA score, we evaluated the feature importance of the RF algorithm. Furthermore, we compared the performances of 4 supervised learning algorithms based on different subsets of features: the first 14 parameters (p14), all 21 parameters (p21), the parameters without performance scales and ICU and hospital stays (p18), and their combinations with the inclusion of the calculated IDDA score (+score) and demographic information (+demo; see Table E1). When the IDDA scores of patients with CVID were compared to those of patients with any other clinically or genetically defined IEI/PID, age and IGRT score stood out as most relevant features to discriminate CVID from the rest of the diagnoses, followed by the calculated IDDA score total (parameters "age," "idda_score"; "ig_score"; Fig 2, *A, left* and *right*), whereas most other features were relatively unspecific for CVID. When only 18 IDDA parameters were chosen as input, IGRT—that is, whether a patient received IGRT or not—was identified to yield the highest MDI and thus was the most distinctive feature of CVID compared to other diagnoses (Fig 2, *A, right;* 18 IDDA parameters). Fig 2, *B-G,* shows the IDDA feature importance plots of exemplary IEI/PID with immune dysregulation. Although derived from relatively small patient subcohorts and calculated with AI, they largely confirm the distinctive pattern of manifestations specific

TABLE I. Patient cohort and diagnosis distribution of IEI/PID patients with features of immune dysregulation with IDDA score data registry entries

Characteristic	Untreated	Total
Individual patients	725	825
Sex distribution (female)	330	368
Age (years), median (range)	23.5 (0.2-84.2)	22.7 (0.2-84.2)
All patient visits (including follow-up)	913	1043
IEI disease entities total (≥10 per disease)	83 (17)	89 (24)
IUIS category*		
Primary antibody deficiencies	350	375
CID with associated or syndromic features	124	134
Diseases of immune dysregulation	64	92
Combined immunodeficiencies	71	84
Phagocytic disorders	50	62
Autoinflammatory disorders	21	31
Complement deficiencies	26	28
Defects in innate immunity	17	17
Bone marrow failure	2	2
IEI diagnoses with ≥5 patients*		
CVID	182	196
APDS	30	37
DGS	34	36
HIES	33	35
Agammaglobulinemia	30	35
CID	27	31
CGD	22	30
Unclassified antibody deficiency	27	27
ATM	26	27
IgG subclass deficiency	26	26
Early-onset multiorgan autoimmunity	18	26
IgA deficiency	24	24
Severe CID	19	22
ALPS	14	18
Congenital neutropenia	14	14
NBS1	7	10
CSR/HIGM (hyper-IgM)	8	10
CMC	9	9
FMF	7	9
Other autoinflammatory (known gene)	7	9
NFKB1 deficiency	8	8
APECED	5	7
FHLH	5	7
HAE (C1Inh)	7	7
LRBA deficiency	4	7
C2 deficiency	6	6
IgA with IgG subclass deficiency	6	6
WAS	5	6
TLR/NF-κB	6	6
XLP	5	6
Netherton syndrome	4	5
Complement ID, unclassified	5	5
Type 1 interferonopathies	2	5
MSMD	2	5
Genetic defects with ≥5 patients*		
Del 22q11.2	33	34
PIK3CD (PI3K-delta)	23	27
ATM	26	27
BTK	23	26
CTLA-4	9	15
ALPS-FAS	13	15
TACI	12	13
GP91-phox (*CYBB*)	9	13
PIK3R1	7	10
STAT3-LOF	10	10
NBS1	7	10
STAT3-DN	9	10
DOCK8	8	9
MEFV	7	9
NFKB1	8	8
WAS	6	7
AIRE	5	7
STAT3-GOF	5	7
LRBA	4	7
DCLRE1C (*Artemis*)	6	7
C2	6	6
RAG1	4	6
STAT1	5	5
IKBKB	5	5
RAG2	4	5
p47-phox (*NCF1*)	5	5

Data are shown as numbers unless otherwise indicated. Data were collected from April 2022 to November 2024 by participants of ESID registry (physicians and documentarists) in 84 centers across Europe and neighboring countries. Patients who had undergone prior definitive therapy were excluded. *Untreated* refers to immune-modifying (eg, immunosuppressive and anti-inflammatory) treatment, excluding immunoglobulin replacement or anti-infective therapy, within last 3 months.
AIRE, Autoimmune regulator; *APECED,* autoimmune polyendocrinopathy candidiasis ectodermal dystrophy; *ATM,* ataxia-telangiectasia mutated; *CGD,* chronic granulomatous disease; *CID,* combined immunodeficiency disease; *CMC,* chronic mucocutaneous candidiasis; *CSR/HIGM,* class-switch recombination/hyper-IgM; *CYBB,* cytochrome *b* subunit beta; *DN,* double negative; *DOCK8,* dedicator of cytokinesis protein 8; *FAS,* FAS receptor; *FHLH,* familial hemophagocytic lymphohistiocytosis; *FMF,* familial Mediterranean fever; *GOF,* gain of function; *HAE (C1INH),* hereditary angioedema with impaired C1 inhibitor; *IKBKB,* inhibitor of nuclear factor kappa B kinase subunit beta; *LOS,* loss of function; *MEFV,* mutations in Mediterranean fever; *MSMD,* Mendelian susceptibility to mycobacterial disease; *NCF1,* neutrophil cytosolic factor 1; *PIK3CD,* phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit delta; *PIK3R1,* phosphatidylinositol 3-kinase phosphorylate; *RAG1/2,* recombinase activating gene 1/2; *STAT1/2,* signal transducer and activator of transcription 1/2; *TACI,* transmembrane activator and CAML interactor; *TLR/NF-κB,* Toll-like receptor/nuclear factor kappa–light-chain enhancer of activated B cells; *WAS,* Wiskott-Aldrich syndrome; *XLP,* X-linked lymphoproliferative disease.
*Numbers are for individual patients, not their sum of patient visits.

each IEI/PID applied by human intelligence. This calculation was performed for each of the diagnoses against the others (one vs the rest; see Fig E2 in the Online Repository available at www.jacionline.org), and including or omitting treated patients did not make a difference (data not shown). To verify the results of RF-based feature importance calculations and interpret individual predictions, we performed the LIME technique in selected cases, which overall well conformed with the ranking of detected feature importance (Fig E2, *B*). When we performed IDDA feature importance analyses across all patients with clinically or genetically defined diagnoses, or across the 9 present IUIS categories of IEI/PID in our entire dataset, the IDDA score total was the most distinctive feature after age (Fig 3). Here, the MDI measures how much each feature reduced uncertainty on average when used to split data via a decision tree, whereas the error bars represent the standard deviation of the MDI values across cross-validation folds during a grid search. We note that the model's

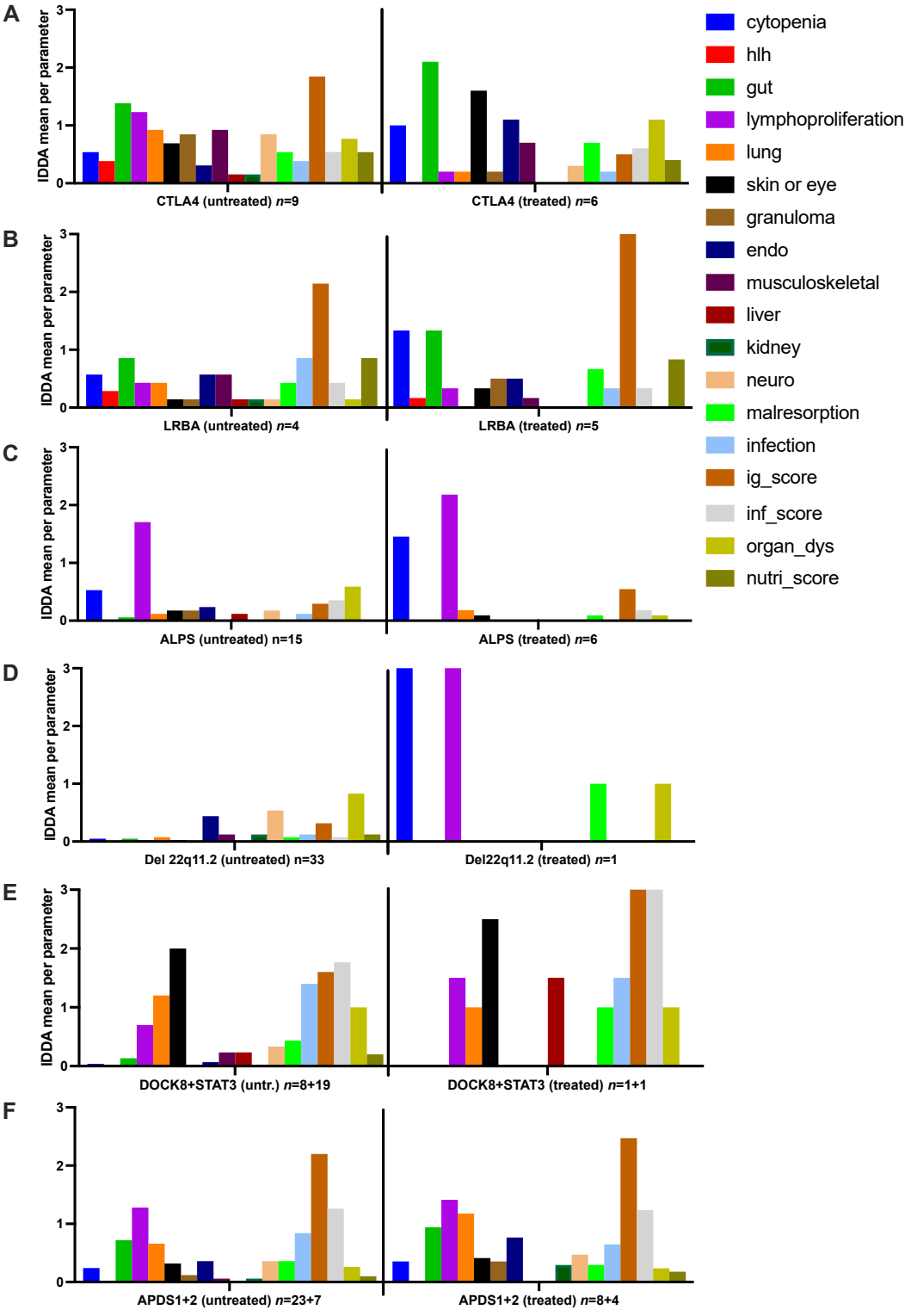


FIG 1. Phenotype profiles based on IDDA score parameters of IEI/PID patients with immune dysregulation. Plots visualize pattern recognition by human intelligence taking place at patient visits, where organ involvement and disease burden are semiquantitatively scored. Abbreviated IDDA score parameters are as outlined in Table E1. Subset of patients was analyzed after or with ongoing immune-modifying therapy within last 3 months ("treated," *right*). IDDA means per parameter of small patient subcohorts are shown on y-axis for CTLA-4 haploinsufficiency **(A)**, LRBA deficiency **(B)**, ALPS **(C)**, DGS **(D)**, HIESs due to *DOCK8* or *STAT3* mutations **(E)**, and APDS types 1 and 2 **(F)**. *DOCK8,* Dedicator of cytokinesis protein 8; *STAT3,* signal transducer and activator of transcription 3. Designed with GraphPad Prism v10.4 (GraphPad Software, La Jolla, Calif).

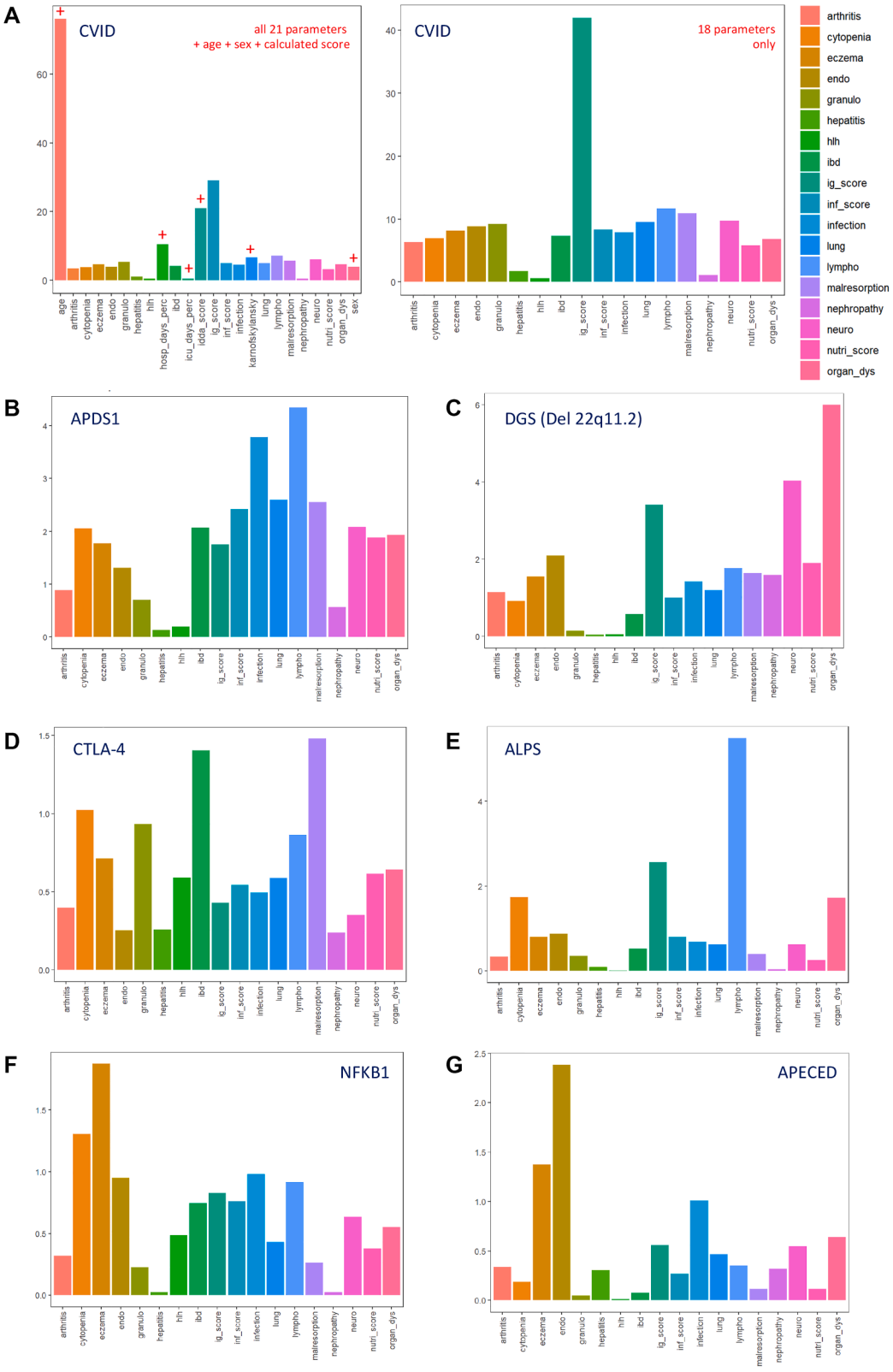


FIG 2. Feature importance analyses of IDDA score parameters in various clinically or genetically defined IEI/PID. Bar diagrams show importance of graded IDDA score disease manifestations for defining specific IEI/PID diagnoses as calculated by RF, as described in Methods section. Abbreviations of IDDA parameters are defined in Table E1 and are listed in alphabetical order. **(A)** Importance of entire set of 21 IDDA score parameters and age, sex, and calculated score sum ("idda_score"; in alphabetical order on x-axis) at *left,* and reduced 18-parameter set only at *right.* **(B-G)** Exemplary feature importances of IDDA scores in patients with different clinically or genetically defined IEI/PID. Entire set of analyses is shown in Fig E2.

performance remained almost identical when the IDDA score total was omitted, thus corroborating the balance of the score calculation with respect to the weighting of individual phenotypic parameters (Fig 3, *C*). Thus, when included in the data input, the IDDA score total was always among the top features without strongly affecting the order of the following features, thereby supporting its potential diagnosis-stratifying power (Fig 3 and Fig E2). To address the inherent phenotypical heterogeneity within the CVID patient cohort, we restricted our ML analyses that aimed to predict monogenic IEIs to only include genetically defined CVID cases (eg, due to variants in *TACI, NFKB2, ICOS, CD19,* or *IRF2BP2*), while clinically diagnosed CVID patients were excluded from these specific tasks. In RF feature importance analysis and UMAP, both genetically and clinically diagnosed CVID cases were included because our goal was to explore broader phenotypic relationships among IEI/PID patients rather than predict monogenic diagnoses.

IDDA score–based diagnosis prediction approaches

On the basis of the distinctive phenotype profiles and feature importances of the collected IDDA datasets from IEI/PID patients, we next sought to classify their diagnoses by using unsupervised learning approaches. We performed UMAP to identify clusters of patients' IDDA score (18 parameters) datasets and labeled them according to their genetic or clinical diagnosis ("gene_pid"). All these analyses were conducted using only a few selected, genetically defined disease subsets to facilitate cluster detection, except in a few instances where CVID datasets were included to explore whether clustering could still be observed despite their anticipated heterogeneity. Fig 4, *A,* shows UMAP plots for ALPS and HIES patients at their first visit, without documented immune-modifying therapy in the last 3 months ("first visit, untreated"), where a larger number of datasets was plotted for the same comparison (ALPS vs HIES) when data from all their visits, including treated patients, were plotted (Fig 4, *B*). We performed the same analysis to compare CTLA-4 haploinsufficiency, LRBA deficiency, and DGS (Fig 4, *C*), agammaglobulinemia, BTK deficiency, DGS, and APDS types 1 and 2 (Fig 4, *D*), AT and Nijmegen breakage syndrome (NBS) with agammaglobulinemia and BTK deficiency (Fig 4, *E*), and AT, NBS, HIES, and various phagocytic disorders (Fig 4, *F*). When we included CVID, we found, as expected, that patients with CVID (at 210 visits of 196 patients) did not show one specific cluster but were dispersed over a few subclusters and overlapped in part with more distinct clusters of patients with NFKB1 or -2 deficiencies, with AT and NBS, or with BTK deficiency, agammaglobulinemia, DGS, or APDS types 1 and 2 (Fig 4, *G* and *H*). These data show that registry-derived IDDA score datasets of patients with immune dysregulation allow us, at least to some degree, to classify a diagnosis with ML assistance based on structured and semiquantitatively graded phenotypic criteria. To provide the entire dataset and permit selection or deselection of any IEI/PID for clustering on the basis of IDDA scores, we created an interactive figure with all patients' IDDA scores in a comprehensive UMAP plot, which is available in the web application in the Online Repository.

The ultimate goal of this study was to learn about phenotype profiles from partially genetically defined IEI/PID and train models to predict the diagnosis of undiagnosed patients, or to identify missing and expectable phenotypic features in diagnosed patients. We therefore compared the performance of supervised learning algorithms for different hierarchies (level of disease classification) of a disease from more general (IUIS categories) to very specific (monogenetic IEIs) and created a label for a preselected group of characteristic IEIs, or cardinal IEI/PID. The balanced accuracy for the prediction of a class was 44% for the IUIS categories, 73% for monogenetic defects, 78% for any underlying monogenic disorder, 57% for the cardinal IEI/PID, and 40% for the combined genetic defects and nongenetically diagnosed IEI/PID (Table II). As a clinical approach to increase the prediction accuracy while still reducing the number of likely differential diagnoses, we tested a top 3 prediction, where the 3 most likely classes (diseases) were proposed as possible candidates when diagnosing a patient's disease. We observed a top 3 accuracy of 86% for the IUIS categories, up to 93% for monogenetic defects, 87% for the cardinal IEI/PID, and 81% for combined genetic defects and nongenetically diagnosed IEI/PID (Table II). After performing different cross-validation approaches, we observed a higher accuracy for a higher number of cross-validation folds (ratio of training vs test datasets) when a minimum number of 10 patients was required (93% top 3 accuracy, 8 diseases predicted) compared to lower folds when at least 4 patients were included (62% top 3 accuracy, 34 diseases predicted; Table II).

DISCUSSION

To our knowledge, this is the first ESID registry substudy to apply supervised and unsupervised ML for disease classification in IEI/PID. The 21 graded parameters of the IDDA2.1 score—reflecting organ involvement and disease activity with a focus on immune dysregulation—served as semiquantitative prestructured input measures for ML. Although patient numbers were small in most disease subcohorts, >1,000 unbiased entries of IDDA2.1 scores from physicians and documentarists in 84 centers across Europe were sufficient to demonstrate proof of principle that a small subset of patient phenotype registry data can effectively predict IEI diagnoses, genetic underpinnings, and expected disease manifestations. In contrast to approaches where data from electronic health records were extracted and then shown to be able to predict the probability of rare diseases like an IEI/PID in general or of CVID or APDS specifically,[22,49-52] we here demonstrate the possibility of clustering different IEI/PID according to their phenotype profiles, thus opening new avenues for using a rare-disease patient registry as a clinical management tool in the future.

Limitations and challenges

Data size. Despite international collaboration, working with rare immunodeficiencies inevitably involves small sample sizes. This limits prediction accuracy, leads to lower stability of ML, increased risk of overfitting, and lower reliability of the test series used to validate the algorithms; they can be highly sensitive to noise and outliers in the data. The appropriate choice of the number of cross-validation folds as well as clinically relevant pretests and the formation of super groups play an important role. Furthermore, an exchange with other disciplines that work with small samples, such as neuroscience (brain imaging studies),[53] astrophysics (rare cosmic events or phenomena),[54] forensics (evidence analysis with limited case data),[55] and materials science

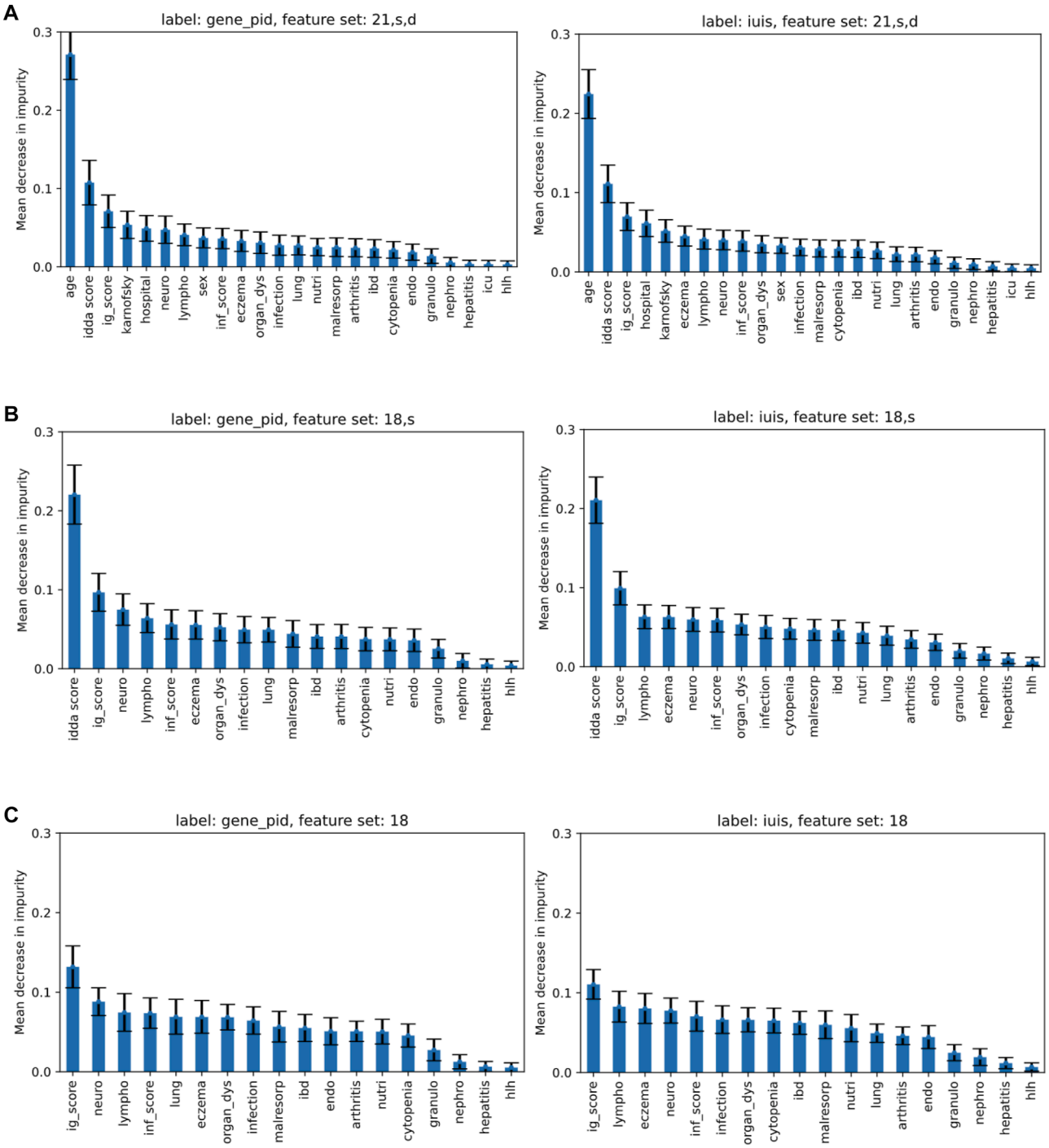


FIG 3. IDDA feature importances (FI) ranked by MDI. FI across all patients with clinically or genetically defined diagnoses *(left),* or across 9 present IUIS categories of IEI/PID *(right).* Calculated IDDA score total was most distinctive feature after age, shown for parameter groups of 21 parameters + score **(A)**, 18 parameters + score **(B)**, or 18 parameters without score **(C)**. Abbreviations of IDDA parameters are defined in Table E1.

(new or rare materials with limited test series),[56] can provide ideas on the use of nonstandard methods and might promote the development of new ML methods.

Human bias. Clinical grading scales introduce subjectivity, with assessments varying across physicians. Moreover, the availability of genetic diagnoses depends on current research and diagnostic access. Missing data may reflect diagnostic limitations rather than the absence of genetic defects. Thus, ML models must be interpreted with caution, particularly regarding monogenic disease attribution.

Label ambiguity. Longitudinal data raise challenges in distinguishing new disease manifestations from evolving

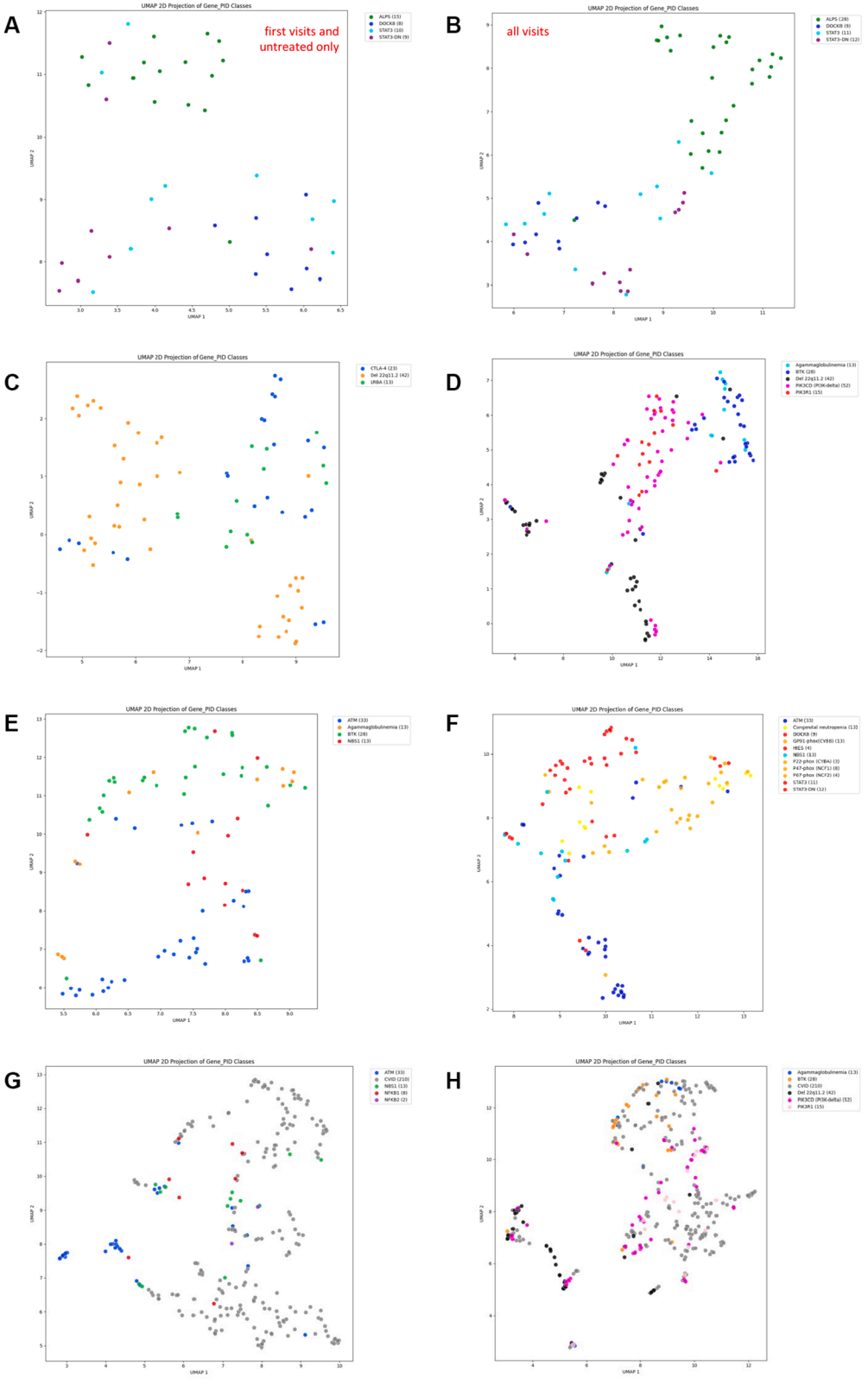


FIG 4. UMAP plots with clusters of IDDA score profiles of patients with selected IEI/PID. (**A** and **B**) IDDA score profile clusters of patients with ALPS or HIES (STAT3 is short for STAT3-LOF); patient visit numbers are shown in parentheses for untreated patients at their first visits *(A)* or for all patients and all visits **(B)**. **(C-H)** UMAP plots for all patients and all their visits, and various selected disease combinations of patients with AT, agammaglobulinemia, BTK deficiency, APDS types 1 and 2, DGS, CTLA-4 haploinsufficiency, LRBA deficiency, congenital neutropenia, chronic granulomatous disease, Nijmegen breakage syndrome, NFKB1 or NFKB2 deficiency, and CVID, as indicated. Web application in Online Repository is interactive UMAP plot where all patient data are provided; data may be selected or deselected to enable testing any combination. *STAT3-LOF,* Signal transducer and activator of transcription 3 loss of function.

TABLE II. Accuracy of diagnosis prediction solely based on IDDA parameters

Prediction	No. of classes in cohort	No. of patients [train | test]*	Minimum patients per class	No. of classes predicted	Balanced with top 1 accuracy	Top 3 accuracy
Genetic defect	114	412 [309 | 103]	10	8	69.7	93.2
			4	34	21.8	62.1
IUIS category	9	723 [542 | 181]	10	8	43.5	85.6
Gene or IEI/PID	137	725 [543 | 182]	10	12	40.1	81.3
Cardinal IEI (25 genes)	12	168 [180 | 61]	6	11	59.4	86.9

*Number reflects only first visits of patients. Balanced accuracy is average of recall (sensitivity) across all classes.

presentations of the same disease. Additionally, treatment timing and response significantly affect observed severity, which may confound model training if not properly accounted for.

Methodologic constraints. Because the number of classes to be predicted varies depending on the number of folds selected for cross-validation, the accuracy changes. Fewer folds mean that classes with fewer patients can be included in the prediction, but this also means that less training data are available per class and the prediction may be less accurate. Because the ESID-IDDA database contains a number of classes with sample sizes that do not meet the cross-validation criterion (with fewer patients than the number of folds), these classes must either be neglected or combined into a supergroup, or a joint, nonspecific group. Such a supergroup combination introduces an arbitrary group of diseases without a specific relationship that could be included in the training. In this case, unknown patients would either be specifically predicted into the classes with the required minimum number of patients or into the nonspecific supergroup. In the latter, no pretest (other than an assumption of monogenetic IEIs) would be required.

Model optimization. The logistic regression model did not consistently converge, suggesting that further refinement in transforming discrete features could enhance its performance. While the study utilized selected supervised and unsupervised learning techniques for prediction and clustering, there remains potential to explore a broader range of methods. Additionally, future research could benefit from incorporating more clinical measures; this study primarily focused on already established parameters. Given that the parameters graded from 0 to IV are ordinal, using the median may provide a more accurate representation of phenotype profiles. The prediction accuracy may be improved by analyzing the confusion matrices of each model to find out which classes are predicted to be better or worse, especially with regard to the top 3 approach.

Outlook. Similarities in phenotypes are likely to be due to similar or identical activated pathophysiologic and metabolic pathways. These promise the possibility of a new, additional label on which ML could be trained. Furthermore, information about the expression of genetic defects or of other -omics data (eg, proteomics, transcriptomics) could be added to the dataset and helpful for improved accuracy in predictions by ML. Moreover, for further projects such as prediction of treatment response, clear information on the clinical presentation before the start of treatment, time of treatment start, and change in clinical presentation during follow-up is essential. Ultimately, in patients who lack a genetic diagnosis, IDDA-based clustering could support phenotype-targeted treatment decisions or clinical trial enrollment in the future.

In order to improve the evaluation of the response to a therapeutic intervention, an updated IDDA2.2 version was made in July 2024 with slight modifications of the definitions in the grading of the parameters (see Table E2 in the Online Repository available at www.jacionline.org). Accordingly, a parameter can be rated with a lower grade (2°) if there is a response to continuous treatment while still receiving treatment. This IDDA2.2 score is already being used in an ongoing large follow-up study of the Inborn Errors Working Party of the European Society for Blood and Marrow Transplantation after Fox et al[57] detected a significant association of low pretransplant IDDA scores with poor outcomes in adolescents and adults with IEI/PID. In addition, to also be applicable to interventional (drug) trials, another version with a slightly extended grading definition was developed (IDDA2.2CT; Table E1). Subsequently, AI-assisted extraction of IDDA score information from electronic health records could be analyzed on the basis of results from these studies and could learn continuously, thereby improving its potential future application as a clinical guidance tool.

Conclusion

In this study, we aimed to strengthen the evidence supporting the clinical parameters used in the IDDA2.1 score to identify the type and severity of IEI. We hypothesized that registry data derived from IEI/PID patients, particularly the structured and semiquantitative IDDA2.1 data, represent an ideal foundation for training AI models to classify IEI/PID diagnoses and predict their clinical manifestations. While current predictions substantially outperform random chance, achieving consistent and reliable performance will require larger patient datasets. The top 3 prediction approach demonstrates promising clinical applicability, but it still needs refinement in terms of sensitivity and specificity for different medical contexts. As with many rare diseases, small sample sizes inherently limit model robustness.

Feature importance analysis further suggests a potential hierarchy of clinical parameters, offering guidance on which features may be most critical for diagnosis and follow-up. These insights may inform the development of a future IDDA3.0 score, potentially incorporating additional easily accessible clinical variables. Dimensionality reduction techniques such as UMAP proved effective for visualizing high-dimensional data, clustering phenotypically similar patients, and revealing intrinsic patterns. This visualization approach, alongside phenotype profiling tools like kaleidoscope or pipe plots, enhances interpretability for clinicians and supports decision-making in diagnosis, treatment selection, and response evaluation—even without the need for statistical expertise. Future studies could include follow-up

patient groups and integrate treatment data to explore the feasibility of predicting disease progression and therapy response using the IDDA score framework.

DISCLOSURE STATEMENT

The programming of the IDDA2.1 score module in the ESID registry was financially supported by a grant from Takeda, Austria, who had no role in the study design, study conduct, data collection, data management, data analysis, data interpretation, or writing of this report. M.G.S. was in part funded by the Styrian Children's Cancer Aid Foundation (Steirische Kinderkrebshilfe).

Declaration of generative AI in scientific writing: The authors used ChatGPT for partial text editing to improve readability; they then reviewed and edited the text and take full responsibility for the content.

Disclosure of potential conflict of interest: M. G. Seidel received advisory board honoraria from Pharming and Takeda unrelated to this work. The rest of the authors declare that they have no relevant conflicts of interest.

Members of the ESID Registry Working Party: Abd Elaziz, Dalia; Abdelkader, Sohilla Lofty M.; Abitbol, Avigaelle; Abolhassani, Hassan; Abrahim, Lalash; Abuzakouk, Mohamed; Accardo, Pietro Andrea; Achir Moussouni, Nabila; Afonso, Veronica; Agyeman, Philipp; Ahlmann, Martina; Aiuti, Alessandro; Akl, Abla; Aksu, Güzide; Albers, Kim; Albert, Michael H.; Alecsandru, Meda Diana; Aleinikova, Olga; Aleshkevich, Svetlana; Alkady, Radwa Salah Eldeen Youssif; Allende, Luis; Alligon, Mickaël; Allwood, Zoe; Alsina, Manrique de Lara Laia; Ambrosch-Barsoumian, Daniela; Ameshofer, Lisa; Amour, Kenza; Anadol, Evrim; Ananthachagaran, Ariharan; Andriamanga, Chantal; Andris, Julia; Andritschke, Karin; Angelini, Federica; Ankermann, Tobias; Apel, Katrin; Arami, Siamak; Ardeniz, Ömür; Arkwright, Peter; Arnold, Karina; Ascherl, Rudolf; Assam, Najla; Assia-Batzir, Nurit; Atschekzei, Faranaz; Aumann, Sybille; Aumann, Volker; Aurivillius, Magnus; Ausserer, Bernd; Avcin, Tadej; Aydemir, Sezin; Aygören-Pürsün, Emel; AyvazDeniz, Cagdas; Azzari, Chiara; Babacar, Lo; Bach, Perrine; Bachmann, Sophie; Bader, Peter; Bakhtiar, Shahrzad; Bancé, Renate; Bangs, Catherine; Barfusz, Katerina; Baris, Safa; BarlanIsil, B; Bartsch, Michaela; Baselli, Lucia Augusta; Batlle-Maso, Laura; Baumann, Helge; Baumann, Ulrich; Baumeister, Veronika; Baxendale, Helen; Bazen, Suzanne; Beaurain, Beatrice; Beauté, Julien; Bechar-Makhloufi, Mounia; Beck, Norbert; Becker, Brigitta; Becker, Christian; Behrends, Uta; Beider, Renata; Beier, Rita; Belkacem, Amel; Belke, Luisa; Bellert, Sven; Belohradsky, Bernd H.; Ben-Bouzid, Aouatef; Benoît, Vincent; BenSlama, Lilia; Berdous-Sahed, Thamila; Bergils, Jan; Berglöf, Anna; Bergman, Peter; Bernat-Sitarz, Katarzyna; Bernatoniene, Jolanta; Bernatowska, Ewa; Bernbeck, Benedikt; Bertolini, Elena; Bethune, Claire; Beuckmann, Kai; Bhole, Malini; Biegner, Anika-Kerstin; Bielack, Stefan; Bienemann, Kirsten; Bigl, Arndt; Bigorgne, Amélie; Bijl, Marc; Binder, Nadine; Bitzenhofer-Grüber, Michaela; Blanchard Rohner , Geraldine; Blank, Dagmar; Blattmann, Claudia; Blau, Julia; Blaziene, Audra; Blazina, Stefan; Bloomfield, Markéta; Blume, Roswitha; Boardman, Barbara; Bode, Sebastian; Boelens, Jaap-Jan; Boesecke, Christoph; Bogaert, Delfien; Bogner, Johannes; Bohynikova, Nadezda; Booth, Claire; Bordon, Victoria; Borkhardt, Arndt; Börries, Melanie; Borte, Michael; Borte, Stephan; Bossaller, Lukas; Bossard, Madeleine; Botros, Jeannet; Boucherit, Soraya; Boutros, Jeannette; Boyarchuk, Oksana; Boyman, Onur; Boztug, Kaan; Branco Pereira da Silva, Sara; Braschler, Thomas; Bravo, Sophie; Bredius, Robbert; Bright, Philip; Brito de Azevedo Amaral , Carolina; Brodszki, Nicholas; Brodt, Grit; Brolund, Allan; Brosselin, Pauline; Brummel, Bastian; Brun-Schmid, Sonja; Brunner, Jürgen; Bruns, Roswitha; Buchta, Christina; Buck, Dietke; Bücker, Aileen; Buckland, Matthew; Bührlen, Martina; Burdach, Stefan; Burns, Siobhan; Burton, Janet; Caminal, Luis; Cancrini, Caterina; Candotti, Fabio; Canessa, Clementina; Cant, Andrew J; Cantoni, Nathan; Capilna, Brindusa; Caracseghi, Fabiola; Caragol, Isabel; Carbone, Javier; Carrabba, Maria; Casanova, Jean-Laurent; Chamberlain, Latanya; Chandra, Anita; Chantrain, Christophe; Chapel, Helen; Chassot, Julie; Chee, Ronnie; Chinello, Matteo; Chopra, Charu; Chovancova, Zita; Christmann, Martin; Chrzanowska, Krystyna; Ciznar, Peter; Claes, Karlien; Classen, Carl Friedrich; Classen, Martin; Cochino, Alexis-Virgil; Condliffe, Alison M.; Corbacioglu, Selim; Cordeiro, Ana Isabel; Cordier Wynar, Donatienne; Core, Claire; Costes, Laurence; Coulter, Tanya; Courteille, Virginie; Cristina, Maria; Cucuruz, Maria; Dabrowska, Leonik Nel; Dabrowska Leonik, Nel; Dähling, Mandy; Daly, Mary Louise; Daniel, Claudia-Sabrina; Danieli, Maria Giovanna; Darroch, James; Davies, Graham; De Gracia Roldan, Javier; de Nadai, Narimene; de Schutter, Iris; De Vergnes, Nathalie; de Vries, Esther; de Witte, Josine; deBaets, Frans; Debert, Theo; DeBoeck, Christiane; Defila, Corina; Deimel, Judith; Delaplace, Diane; Dellepiane, Rosa Maria; Dellert, Nelli; Delliera, Laura; Delor, Anita; Demel, Ulrike; Dempster, John; den Os, M.M.; Dengg, Rosmarie; Desisa, Sora Asfaw; Desta, Alexandra; Detkova, Drahomira; Dewerchin, Maite; Dieli Crimi, Romina; Dilloo, Dagmar; Dimitriou, Florentia; Dinges, Sarah Svenja; Dinser, Jasmin; Dipani, Nabila; Dirks, Johannes; Dittrich, Anna-Maria; Djermane, Lylia; Dogru, Yagmur; Dogu Esin, Figen; Dombrowski, Angelika; Dominguez Escobar, Julia; Döring, Michaela; Drabe, Camilla Heldbjerg; Drerup, Susann; Drexel, Barbara; Driessen, G.J.A.; Dückers, Gregor; Dudoit, Yasmine; Duppenthaler, Andrea; Ebetsberger-Dachs, Georg; Ecser, Mate Barnabas; Edgar, J. David; Eekman, Maartje; Ehl, Stephan; Ehrat, Rosanna; Eisl, Eva; Ekwall, Olov; El Hawary, Rabab; El-Helou, Sabine M.; El-Marsafy, Aisha; Elbe, Sarina; Elcombe, Suzanne; Eldash, Alia; Eldeen, Youssif Alkady Radwa Salah; Ellerbroek, P.M.; Elling, Roland; Elliott, Jane; Elmarsafy, Aisha; Engelhardt, Angelika; Ernst, Diana; Ersoy, Fügen; Esper, Stefanie; Esteves, Isabel; Etzioni, Amos; Exley, Andrew; Faber, Martin; Fabio, Giovanna; Fahrni, Gaby; Faletti, Laura Eva; Faletti, Laura Eva; Farber, Claire-Michele; Farela Neves, João; Faria, Emilia; Farkas, Henriette; Farmaki, Evangelia; Faßhauer, Maria; Fasth, Anders; Fätkenheuer, Gerd; Faustmann, Stefanie; Fecker, Gisela; Feighery, Conleth; Feiterna-Sperling, Cornelia; Fernandez-Cruz, Perez Eduardo; Ferreira Concalo, Cordeiro; Ferster, Alice; Feuchtinger, Tobias; Feyen, Oliver; Finke, Daniela; Fischer, Alain; Fitter, Sigrid; Flaschberger, Stefan; Fleckenstein, Lucia; Föll, Dirk; Fontana, Adriano; Forino, Concetta; Förster-Waldl, Elisabeth; Franchet, Nora; Freitag, Dagmar; Frey, Urs P.; Frick, Hannah Margarete; Friedel, Elisabeth; Friedrich, Wilhelm; Frisch, Barbara; Frischknecht, Lukas; Fritzemeyer, Stephanie; Gagro, Alenka; Gahr, Manfred; Galal, Nermeen Mouftah; Gambineri, Eleonora; Gamper, Agnes; Gams, Franziska; Ganzow, Astrid; Garcelon, Nicolas; Garcez, Tomaz; GarciaPrat, Marina; Gardulf, Ann; Garibay, Janine; Garwer, Birgit; Gathmann, Jonathan; Gathmann, Benjamin; Gebauer, Corinna; Geberzahn, Linda; Geikowski, Tilman; Geisen, Ulf; Gemander, Christiane; Gennery, Andrew R.; Gerisch, Marie; Gernert, Michael; Gerrer, Katrin; Gerschmann, Stev; Ghosh, Sujal; Giannini, Carolin; Gil Herrera, Juana; Gimenez Sanz, Noemi; Girndt, Matthias; Girrbach, Ramona; Girschick, Hermann; Gkantaras, Antonios; Gkougkourelas, Ioannis; Gładysz, Dominika; Gnatowski, Susanne; Gnodtke, Elisabeth; Goda, Vera; Goddard, Sarah; Goebel, Daniela; Goffard, Jean-Christophe; Goldacker, Sigune; Gollowitsch, Eva Maria; Gomes, Manuella; Gompels, Mark; González, Míriam; Gonzalez Granado, Luis Ignacio; Gordins, Pavels; Gößling, Katharina; Göschl, Lisa; Gossens, Lucy; Gowin, Ewelina; Graafen, Lea; Graca, Leo; Gradauskiene-Sitkauskiene, Brigita; Graf, Dagmar; Graf, Norbert; Grange, Elliot; Grashoff, H.Anne; Greil, Johann; Grigoriadou, Sofia; Grimbacher, Bodo; Gronlund, Helen; Groß-Wieltsch, Ute; Guerra, Teresa; Guevara-Hoyer, Kissy; Gueye, Mor Seny; GueyeMor, Seny; Guibert, Noemie; Gülnur, Birgit; Güngör, Tayfun; Guseva, Marina; Guzman, David; Haag, Marcel; Haase, Gabriele; Haenicke, Henriette; Haerynck, Filomeen; Hafsa, Ines; Hagin, David; Haliti, Emine; Hallek, Michael; Hammarstroem, Lennart; Hancioglu, Gonca; Handgretinger, Rupert; Hanitsch, Leif G.; Hansen, Susanne; Hariyan, Tetyana; Harrer, Thomas; Hassunah, Pia; Hatzistilianou, Maria; Hauck, Fabian; Hauser, Thomas; Haverkamp, Margje H.; Hayman, Grant; Heath, Paul; Hedrich, Christian; Heeg, Maximilian; Heike, Michael; Heimbrodt, Martin; Heine, Sabine; Heininger, Ulrich; Heinrich, Christian; Heinz, Valerie; Heitger,

Andreas; Helbert, Matthew; Helbling, Arthur; Heldbjerg Drabe, Camilla; Hellige, Antje; Hempel, Julya; Henderson, Karen; Henes, Jörg; Henneke, Philipp; Hennig, Christian; Henrichs, Karin; Herbst, Martin; Hermann, Walter; Hernandez, Manuel; Hernández, Anja; Heropolitanska-Pliszka, Edyta; Herriot, Richard; Herrmann, Friedrich; Herwadkar, Archana; Hess, Christoph; Hess, Ursula; Hesse, Sebastian; Higgins, Sonja; Hilfanova, Anna; Hilpert, Sophie; Hintze, Chantal; Hlaváčková, Eva; Hodl, Isabel; Hodzic, Adna; Hoernes, Miriam; Hoffmann, Christina; Holbro, Andreas; Höllinger, Christiane; Holtsch, Lisa; Holzer, Ursula; Holzinger, Dirk; Hönig, Manfred; Hönscheid, Andrea; Horn, Julia; Horneff, Gerd; Hoyoux, Claire; Hristova, Nataliya; Hübel, Kai; Hübner, Angela; Huemer, Christian; Huissoon, Aarnoud; Hülsmann, Brigitte; Hundsdörfer, Patrick; Huppertz, Hans-Iko; Huß, Kristina; Hussain, Sadia; Husson, Julien; Hüttner-Foehlisch, Tanja; Ijspeert, Hanna; Ikinciogullari, Aydan; İlknur, Kökçü; Irga, Ninela; Jablonka, Alexandra; Jahnz-Rozyk, Karina; Jakob, Marcus; Jakoby, Donate; Jakoby-Gaide, Donate; Jandus, Peter; Jansson, Annette; Jaquet, Melanie; Jardefors, Helene; Jargulinska, Edyta; Jauk, Barbara; Jesenak, Milos; Jilka, Katharina; Jolles, Stephen; Jones, Alison; Jones, Regina; Jonkman-Berk, Birgit; Jönsson, Göran; Joyce, Hilary J.; Juliana, Pricillia; Kabesch, Michael; Kager, Leo; Kahlert, Christian; Kaiser-Labusch, Petra; Kakkas, Ioannis; Kamitz, Dirk; Kanariou, Maria; Kanz, Lothar; Karakoc-Aydiner, Elif; Karanovic, Boris; Kartal-Kaess, Mutlu; Käser, Elisabeth; Katzenstein, Terese L.; Kayserova, Hana; Kelleher, Peter; Kerre, Tessa; Kilic, Sara Sebnem; Kindle, Gerhard; Kirchner, Martina; Kiwit, Simone; Kiykim, Ayca; Klasen, Jessica; Klaudel-Dreszler, Maja; Klein, Ariane; Klein, Christoph; Klein-Franke, Andreas; Kleine, Ilona; Kleinert, Stefan; Klemann, Christian; Klima, Marion; Klocperk, Adam; Kobbe, Robin; Kocacik Uygun, Dilara Fatma; Koch, Melanie; Kochler, Yvonne; Kohistani, Naschla; Kojic, Marina; Kolios, Antonio; Kölsch, Uwe; Koltan, Sylwia; Kondratenko, Irina; Königs, Christoph; Konoplyannikova, Julia; Kopac, Peter; Kopp, Jana; Körholz, Dieter; Körholz, Julia; Korte, Pauline; Kostyuchenko, Larysa; Kötter, Ina; Kracker, Sven; Králícková, Pavlina; Kramm, Christof; Kramme, Philipp; Krausz, Máté; Kreuz, Wolfhart; Krista, Johanna; Kriván, Gergely; Kropshofer, Gabriele; Krüger, Renate; Krystufkova, Olga; Ktistaki, Maria; Kühl, Jörn-Sven; Kühn, Alexander; Kuijpers, Taco W.; Kuis, Wietse; Kullmann, Silke; Kulozik, Andreas; Kumararatne, Dinakantha; Kümmler, Ria; Kündgen, Andrea; Kurenko-Deptuch, Magdalena; KussPaula, Cosima; Kütükcüler, Necil; Lafoix-Mignot, Cécile; Lamers, Beate; Lanbeck, Peter; Landais, Paul; Landwehr-Kenzel, Sybille; Langemeyer, Vanessa; Langer, Thorsten; Lankisch, Petra; Lanz, Nadia; Lara, Manrique de; Lara-Villacanas, Eusebia; Laubenthal, Lisa; Laws, Hans-Jürgen; Leahy, Ronan; Lee, Jae-Yun; Lehmann, Andrea; Lehmberg, Kai; Lehner, Patricia; Leibfrit, Hans; Leistner, Leoni; LeMignot, Loic; Lesch, Petra; Leutner, Simon; Liatsis, Manolis; Libai Véghová, Linda; Liebel, Johanna; Liese, Johannes G.; Linauskiene, Kotryna; Linde, Richard; Linßner, Martina; Lippert, Conrad Ferdinand; Litzman, Jiri; Llobet, Pilar; Lo, Babacar; Lodin, Tariq; Lokaj, Jindrich; Longhino, David; Longhurst, Hilary; Lopes da Silva, Susana; Lorenzen, Catharina; Lougaris, Vassilios; Löw, Doris; Lubatschofski, Annelie; Lucas, Mary; Lutz-Wiegers, Verena; Maaß, Sabine; Maccari, Maria Elena; Macura-Biegun, Anna; Maerz, Vanessa; Maggina, Paraskevi; Mahlaoui, Nizar; Mahrenholz, Hannah; Maier, Sarah; Makhlouf, Mounia; Malfroot, Anne; Malinauskiene, Laura; Mannhardt-Laakmann, Wilma; Manson, Ania; Mantkowski, Felicia; Manzey, Petra; Marasco, Carolina; Marcus-Mandelblit, Nufar; Marg, Wolfgang; Markelj, Gasper; Marodi, Laszlo; Marques, José Goncalo; Marques, Laura; Marschall, Karin; Martínez, Natalia; Martinez de la Ossa Saenz-Lopez, Rafael; Martinez-Saguer, Inmaculada; Martire, Baldassarre; Marzollo, Antonio; Masekela, Refiloe; Masjosthusmann, Katja; Masmas, Tania Nicole; Matamoros, Nuria; Mattern, Jutta; Mau-Asam, Pearl; McDermott, Elizabeth; McIntosh, Nichole; Meglic, Karmen Mesko; Meijer, Ruben; Meinhardt, Andrea; Meshaal, Safa; Messaoud, Yasmina; Meyer, Björn; Meyer-Olson, Dirk; Meyts, Isabelle; Micol, Romain; Micoloc, Bozena; Mielke, Gudrun; Milito, Cinzia; Milota, Tomas; Misbah, Siraj; Mödden, Carolin; Mohr, Michael; Mohrmann, Karina; Moin, Mostafa; Molinos, Luis; Möller, Jana; Möller-Nehring, Sarah; Morbach, Henner; Moschese, Viviana; Moser, Olga; Moshous, Despina; Motkowski, Radoslaw; Motwani, Jayashree; MouftahGalal, Nermeen; Müglich, Carmen; Mukhina, Anna; Muller, Eva; Müller, Christiane; Müller, Gabriele; Müller, Hedi; Müller, Ingo; Müller, Thomas; Müller, Zoe; Müller-Ladner, Ulf; Müller-Stöver, Sarah; Münstermann, Esther; Murtra Garrell, Núria; Muschaweck, Moritz; Mutert, Miriam; Nademi, Zohreh; Naik, Paru; Näke, Andrea; Nalda, Andrea Martin; Nasrullayeva, Gulnara; Naumann-Bartsch, Nora; Nemitz, Verena; Neth, Olaf; Neubauer, Andreas; Neubert, Jennifer; Neumann, Carla; Niehues, Tim; Niemuth, Mara; Nieters, Alexandra; Nieuwhof, Chris; Nolkemper, Daniela; Noorani, Sadia; Noorlander, Budde Adya; Notarangelo, Luigi D; Notheis, Gundula; Nowatsh, Sanam Amelie; Obenga, Gaelle; Ocak, Suheyla; Oker, Mehmet; Olbrich, Peter; Olipra, Anna; Omran, Heymut; Oommen, Prasad; Opitz, Linda; Orosova, Jaroslava; Oskarsdottir, Solveig; Özsahin, Hülya; Pac, Malgorzata; Pachlopnik-Schmid, Jana; Pandolfi, Franco; Papadopoulou-Alataki, Efimia; Papastamatiou, Theodora; Papatriantafillou-Schmieder, Anna; Parra-Martinez, Alba; Paschenko, Olga; Pašić, Srdjan; Pasnik, Jarek; Patel, Smita; Pavlík, Martin; Paz Artal, Estela; Peeters, Anouk; Pereira da Silva, Sara Branco; Perez-Becker, Ruy; Perez-Guzman, Marc; Pergent, Martine; Perlhagen, Markus; Peter, Hans-Hartmut; Petrić, Marin; Pfreundschuh, Michael; Philippet, Pierre; Picard, Capucine; Pietrucha, Barbara; Pietzsch, Leonora; Pignata, Claudio; Piquer Gibert, Monica; Pirolt, Kerstin; Plebani, Alessandro; Pleguezuelo, Daniel E.; Pollok, Katrin; Pommerening, Helena; Popihn, Daniela; Poplonek, Aleksandra; Popp, Marina; Porta, Fulvio; Portegys, Jan; Posfay-Barbe, Klara; Potjewijd, Judith; Poulheim, Sebastian; Prader, Seraina; Prämassing-Scherzer, Petra; Prelog, Martina; Prevot, Johan; Price, Arthur; Price, Timothy; Proesmans, Marijke; Provot, Johan; Pulvirenti, Federica; Quinti, Isabella; Raab, Anna; Rack, Anita; Raffac, Stefan; Ramos Oviedo, Eduardo; Randrianomenjanahary, Philippe; Ranohavimparany, Anja; Raptaki, Maria; Rashidzadeh, Roonaka; Rathwallner, Margit; Reda, Shereen; Redouane, Nahida; Regateiro, Frederico S.; Reichenbach, Janine; Reimers, Bianca; Reinhardt, Cornelia; Reinhardt, Dirk; Reinprecht, Anne; Reiß, Tamara; Reisli, Ismail; Renner, Eleonore; Rezaei, Nima; Richter, Alex; Richter, Darko; Rieber, Nikolaus Peter; Rieckehr, Nadja; Riedel, Marion; Riescher, Heidi; Rischewski, Johannes; Ristl, Nicole; Ritterbusch, Henrike; Ritz, Tanja; Rivier, Francois; Robinson, Peter; Rockstroh, Jürgen K.; Roesler, Joachim; Rofiah, Himatur; Rogerson, Elizabeth; Rolfes, Elisabeth; Roller, Beate; Romanyshyn, Yaryna; Rondelli, Roberto; Roosens, Fien; Rösen-Wolff, Angela; Rösler, Valentina; Roth, Johannes; Rothoeft, Tobias; Roubertie, Agathe; Rübsam, Gesa; Rusch, Stephan; Rutgers, Abraham; Ryan, Paul; Sach, Gudrun; Sadeghi, Kambis; Sahrbacher, Ulrike; Saidi, Angelika; Sanal Tezcan, Özden; Sanchez-Ramon, Silvia; Santos, Juan Luis; Sargur, Ravishankar; Savchak, Ihor; Savic, Sinisa; Schaaf, Bernhard; Schaefer, Marzena; Schäfe, Christina; Scharbatke, Eva; Schatorje, Ellen; Schauer, Uwe; Scheibenbogen, Carmen; Scheible, Raphael; Scheinecker, Clemens; Schiller, Romana; Schilling, Beatrice; Schilling, Freimut; Schlieben, Steffi; Schmalbach, Thilo; Schmalzing, Marc Thomas; Schmid, Pirmin; Schmidt, Nadine; Schmidt, Reinhold Ernst; Schmitz, Monika; Schneider, Dominik; Schneider, Dominik T.; Schneppenheim, Reinhard; Scholtes, Cathy; Schölvinck, E.H.; Schönberger, Stefan; Schreiber, Stefan; Schrijvers, Rik; Schroll, Andrea; Schruhl, Simone; Schrum, Johanna; Schubert, Ralf; Schuetz, Catharina; Schuh, Sebastian; Schulz, Ansgar; Schulz, Claudia; Schulze, Ilka; Schulze-Koops, Hendrik; Schulze-Sturm, Ulf; Schumacher, Eva-Maria; Schürmann, Elvira; Schürmann, Gesine; Schuster, Volker; Schwaneck, Eva; Schwarz, Klaus; Schwarz, Tobias; Schwarze-Zander, Carolynne; Schweigerer, Lothar; Sediva, Anna; Seebach, Jörg; Seger, Reinhard; Segerer, Florian; Seidel, Markus G.; Selle, Barbara; Seneviratne, Suranjith; Seppänen, Mikko; Shabanaj, Hatidje; Sharapova, Svetlana; Shcherbina, Anna; Shillitoe, Benjamin; Siepelmeyer, Anne; Siepermann, Kathrin; Simon, Anna; Simon, Arne; Simon-Klingenstein, Katja; Simonovic, Marija; Simsen-Baratault, Merlin; Sindram, Elena; Skapenko, Alla; Skarke, Maiken; Skomska-Pawliszak, Malgorzata; Slatter, Mary; Smet, Julie; Smith, C. I. Edvard; Sobh, Ali; Sobik, Bettina; Sogkas, Georgios; Sohm, Michael; Solanich, Xavier; Solanich-Moreno, Xavier; Soler-Palacín, Pere; Sollinger, Franz; Somech, Raz; Sonnenschein, Anja; Soresina, Annarosa; Sornsakrin, Marijke; Soura, Stavrieta; Spaccarotella, Sabrina; Spadaro, Guiseppe; Sparber-Sauer, Monika; Specker, Christof; Speckmann, Carsten; Speidel, Lisa; Speletas, Matthaios; Stachel, Klaus-Daniel; Stadon, Catherine; Stanislas, Aurélie; Stapornwongkul, Cynthia; Staus, Paulina; Steck, Regina; Steele, Cathal; Steffin, Herbert; Steiner, Urs; Steinmann, Sandra; Stevens, Wim; Stiefel, Martina; Stieger, Sarah; Stiehler, Sophie; Stimm, Hermann; Stojanov, Silvia; Stoll, Matthias; Stoppa-Lyonnet, Dominique; Strapatsas, Tobias; Strauß, Gabriele; Streiter, Monika;

Strik-Albers, Riet; Strotmann, Gaby; SuárezCasado, Héctor; Subiza, Jose Luis; Sundin, Mikael; Süß, Birgit; Sutter, Fabienne; Szaflarska, Anna; Szemkus, Monika; Tamary, Hannah; Tantou, Sofia; Tarzi, Michael D.; Taschner, Helga; Tedgard, Ulf; Teixeira, Carla; ten Berge, RJM; Tenbrock, Klaus; Tester, Sabine; Tezcan, Ilhan; Thalguter, Sonja; Thalhammer, Julian; Thoma, Katharina; Thomas, Moira; Thomczyk, Fabian; Thon, Vojtech; Thrasher, Adrian; Tierney, Patricia; Tietsch, Nadine; Tommasini, Alberto; Tönnes, Beate; Tony, Hans-Peter; Trachana, Maria; Trapp, Carmen; Tricas, Lourdes; Trindade Neves, Maria Conceicao; Trischler, Jordis; Ubieto, Hugo; Uelzen, Anett; Uhlmann, Annette; Ullrich, Jan; Ullrich, Kurt; Ünal, Ekrem; Urbanski, Gerhard; Urschel, Simon; Uszynska, Aleksandra; Vacca, Angelo; Vaganov, N. N.; Vagedes, Daniel; Valicevic, Stefanie; Vallelian, Florence; van Beem, Rachel T; van Damme, Charlotte; van de Ven, Annick; van den Berg, J. Merlijn; van der Flier, Michiel; van Dissel, J.T.; van Hagen, P.M.; van Montfrans, J.M.; van Ogtrop, Geoffrey; van Rens, Jacqui; van Riel, Christel A.M.P.; van Royen-Kerkhof, Annet; van Well, G.Th.J.; Vasiliki, Antari; Velbri, Sirje; Vencken, Jo; Ventura, Alessandro; Vermeulen, François; Vermylen, Christiane; Viemann, Dorothee; Viereck, Anja; Villa, Anna; Vincke, Jeroen; Vinnemeier-Laubenthal, Lisa; Vo Thi, Kim Duy; Voeller, Mirjam; Vollbach, Kristina; Volokha, Alla; von Bernuth, Horst; von Bismarck, Philipp; Voss, Rebecca; Voß, Sandra; Vural, Yüksel; Wachuga, Heike; Wagner, Norbert; Wagström, Per; Wahle, Matthias; Wahn, Volker; Wapp, Nadine; Warnatz, Klaus; Warneke, Monika; Warris, Adilia; Wasmuth, Jan-Christian; Wasserfallen, Jean-Blaise; Wawer, Angela; Weber, Manfred; Weemaes, Corrie MR; Wege, Lisa; Wehrle, Julius; Weidinger, Stephan; Weiß, Michael; Weißbarth-Riedel, Elisabeth; Werner, Antje; Wessel, Sara; Westkemper, Marco; Wicher, Monika; Wickmann, Lutz; Wiegert, Sabine; Wiehe, Monique; Wiehler, Katharina; Wiesböck, Lydia; Wiesik-Szewczyk, Ewa; Williams, Anthony; Winkler, Beate; Winkler, Christel; Winkler, Martina; Winkler, Melanie; Wintergerst, Uwe; Wisgrill, Lukas; Witte, Torsten; Wittkowski, Helmut; Wolf, Barbara; Wölke, Sandra; Wolschner, Christina; Wolska-Kusnierz, Beata; Wood, Philip; Workman, Sarita; Worth, Austen; Wortmann, Michaela; Wuillemin, Walter Alfred; Wulffraat, Nico M; Wustrau, Katharina; Wyndham-Thomas, Chloé; Yegin, Olcay; Yildiran, Alisan; Yilmaz, Denise; Young, Patrick; Yucel, Esra; Zečević, Milica; Želimir, Erić; Zepp, Fred; Zetzsche, Klaus; Zeuner, Rainald; Zielen, Stefan; Zimmermann, Martina.

We thank Gergana Georgieva at the ESID registry administrative office and all contributors and supporters not already mentioned in the collaborator list of the ESID Registry Working Party (eg, documentarists, study nurses, and data analysts) for their continuous support.

Key messages

- **ML approaches using IEI/PID patient phenotype data derived from an ESID registry study can predict IEI diagnoses, genetic underpinnings, and expectable features.**
- **Proof of principle is provided of the use of rare-disease patient registry data as an AI-assisted clinical guidance tool in the future.**

REFERENCES

1. Seidel MG, Hauck F. Multilayer concept of autoimmune mechanisms and manifestations in inborn errors of immunity: relevance for precision therapy. J Allergy Clin Immunol 2024;153:615-28.e4.
2. Grimbacher B, Warnatz K, Yong PFK, Korganow AS, Peter HH. The crossroads of autoimmunity and immunodeficiency: lessons from polygenic traits and monogenic defects. J Allergy Clin Immunol 2016;137:3-17.
3. Vaseghi-Shanjani M, Smith KL, Sara RJ, Modi BP, Branch A, Sharma M, et al. Inborn errors of immunity manifesting as atopic disorders. J Allergy Clin Immunol 2021;148:1130-9.
4. Bousfiha A, Moundir A, Tangye SG, Picard C, Jeddane L, Al-Herz W, et al. The 2022 update of IUIS phenotypical classification for human inborn errors of immunity. J Clin Immunol 2022;42:1508-20.
5. Poli MC, Aksentijevich I, Bousfiha A, Cunningham-Rundles C, Hambleton S, Klein C, et al. Human inborn errors of immunity: 2024 update on the classification from the International Union of Immunological Societies Expert Committee. Journal of Human Immunity 2025;1:1. https://doi.org/10.70962/jhi.20250003.
6. Thalhammer J, Kindle G, Nieters A, Rusch S, Seppanen MRJ, Fischer A, et al. Initial presenting manifestations in 16,486 patients with inborn errors of immunity include infections and noninfectious manifestations. J Allergy Clin Immunol 2021;148:1332-41.e5.
7. Kindle G, Alligon M, Albert MH, Buckland M, Edgar JD, Gathmann B, et al. Inborn errors of immunity: manifestation, treatment, and outcome—an ESID registry 1994-2024 report on 30,628 patients. medRxiv April 16, 2025. https://doi.org/10.1101/2025.02.20.25322586.
8. Seidel MG. Rethinking PIDs: why the distinction between primary and secondary immune disorders is more frequently relevant than that between inborn and acquired errors of immunity. J Allergy Clin Immunol 2024;153:1543-5.
9. Seidel MG, Kindle G, Gathmann B, Quinti I, Buckland M, van Montfrans J, et al. The European Society for Immunodeficiencies (ESID) registry working definitions for the clinical diagnosis of inborn errors of immunity. J Allergy Clin Immunol Pract 2019;7:1763-70.
10. Seidel MG, Boztug K, Haas OA. Immune dysregulation syndromes (IPEX, CD27 deficiency, and others): always doomed from the start? J Clin Immunol 2016;36:6-7.
11. Barzaghi F, Amaya Hernandez LC, Neven B, Ricci S, Kucuk ZY, Bleesing JJ, et al. Long-term follow-up of IPEX syndrome patients after different therapeutic strategies: an international multicenter retrospective study. J Allergy Clin Immunol 2018;141:1036-49.e5.
12. Gámez-Díaz L, August D, Stepensky P, Revel-Vilk S, Seidel MG, Noriko M, et al. The extended phenotype of LPS-responsive beige-like anchor protein (LRBA) deficiency. J Allergy Clin Immunol 2016;137:223-30.
13. Akalu YT, Bogunovic D. Inborn errors of immunity: an expanding universe of disease and genetic architecture. Nat Rev Genet 2024;25:184-95.
14. Tesch VK, Abolhassani H, Shadur B, Zobel J, Mareika Y, Sharapova S, et al. Long-term outcome of LRBA deficiency in 76 patients after various treatment modalities as evaluated by the immune deficiency and dysregulation activity (IDDA) score. J Allergy Clin Immunol 2020;145:1452-63.
15. Tsilifis C, Speckmann C, Lum SH, Fox TA, Soler AM, Mozo Y, et al. Hematopoietic stem cell transplantation for CTLA-4 insufficiency across Europe: a European Society for Blood and Marrow Transplantation Inborn Errors Working Party study. J Allergy Clin Immunol 2024;154:1534-44.
16. Seidel MG, Tesch VK, Yang L, Hauck F, Horn AL, Smolle MA, et al. The immune deficiency and dysregulation activity (IDDA2.1 "Kaleidoscope") score and other clinical measures in inborn errors of immunity. J Clin Immunol 2022;42:484-98.
17. Yin J, Ngiam KY, Teo HH. Role of artificial intelligence applications in real-life clinical practice: systematic review. J Med Internet Res 2021;23:e25759.
18. Riviere JG, Soler Palacin P, Butte MJ. Proceedings from the inaugural Artificial Intelligence in Primary Immune Deficiencies (AIPID) conference. J Allergy Clin Immunol 2024;153:637-42.
19. Rider NL, Coffey M, Kurian A, Quinn J, Orange JS, Modell V, et al. A validated artificial intelligence–based pipeline for population-wide primary immunodeficiency screening. J Allergy Clin Immunol 2023;151:272-9.
20. Mayampurath A, Ajith A, Anderson-Smits C, Chang SC, Brouwer E, Johnson J, et al. Early diagnosis of primary immunodeficiency disease using clinical data and machine learning. J Allergy Clin Immunol Pract 2022;10:3002-7.e5.
21. Rivière JG, Cantenys-Saba R, Carot-Sans G, Piera-Jiménez J, Butte MJ, Soler-Palacín P, et al. Current perspectives and challenges of using artificial intelligence in immunodeficiencies. J Allergy Clin Immunol 2025;156:878-88.
22. Johnson R, Stephens AV, Mester R, Knyazev S, Kohn LA, Freund MK, et al. Electronic health record signatures identify undiagnosed patients with common variable immunodeficiency disease. Sci Transl Med 2024;16:eade4510.
23. Faouzi J, Colliot O. Classic machine learning methods. In: Colliot O, editor. Machine learning for brain disorders. New York: Springer Nature; 2023. pp. 25-75. https://doi.org/10.1007/978-1-0716-3195-9.
24. Liu Q, Chen AT, Li R, Yan L, Quan X, Liu X, et al. Development and validation of machine learning models for intraoperative blood transfusion prediction in severe lumbar disc herniation. iScience 2024;27:111106.
25. Tiwari P, Upadhyay D, Pant B, Mohd N. Multiclass classification in machine learning algorithms for disease prediction. In: Luhach AK, Jat DS, Hawari KBG, Gao XZ, Lingras P, editors. Advanced informatics for computing research. Cham (Switzerland): Springer; 2022. pp. 102-11.
26. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33:1-22.
27. Cristianini N, Ricci E. Support vector machines. In: Kao MY, editor. Encyclopedia of algorithms. New York: Springer; 2008. pp. 928-32.
28. Mucherino A, Papajorgji PJ, Pardalos PM. *k*-Nearest neighbor classification. In: Data mining in agriculture. New York: Springer; 2009. pp. 83-106.
29. Breiman L. Random forests. Machine Learning 2001;45:5-32.
30. Yang L, Shami A. On hyperparameter optimization of machine learning algorithms: theory and practice. Neurocomputing 2020;415:295-316.

31. Genuer R, Poggi JM, Tuleau C. Random forests: some methodological insights. arXiv November 21, 2008. https://doi.org/10.48550/arXiv.0811.3619.
32. Wenck S, Creydt M, Hansen J, Gärber F, Fischer M, Seifert S. Opening the random forest black box of the metabolome by the application of surrogate minimal depth. Metabolites 2021;12:5.
33. Polewko-Klim A, Lesiński W, Golińska AK, Mnich K, Siwek M, Rudnicki WR. Sensitivity analysis based on the random forest machine learning algorithm identifies candidate genes for regulation of innate and adaptive immune response of chicken. Poult Sci 2020;99:6341-54.
34. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco (Calif): Association for Computing Machinery; 2016. pp. 1135-44.
35. Palit S, Heuser C, de Almeida GP, Theis FJ, Zielinski CE. Meeting the challenges of high-dimensional single-cell data analysis in immunology. Front Immunol 2019;10:1515.
36. Al-Hashem MA, Alqudah AM, Qananwah Q. Performance evaluation of different machine learning classification algorithms for disease diagnosis. International Journal of E-Health and Medical Communications 2021;12:1-28.
37. McKinney W. Data structures for statistical computing in Python. In: van der Walt S, Millman J, editors. SciPy 2010: proceedings of the 9th Python in Science conference; June 28-July 3, 2010. https://doi.org/10.25080/Majora-92bf1922-012.
38. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020;17:261-72.
39. Foundation R. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing; 2022. Available at: https://www.r-project.org.
40. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature 2020;585:357-62.
41. Wickham H, François R, Henry L, Müller K, Vaughan D. dplyr: a grammar of data manipulation. R package. GitHub, v1.1.4, November 17, 2023. Available at: https://github.com/tidyverse/dplyr/.
42. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. Journal of Machine Learning Research 2011;12:2825-30.
43. Liaw A, Wiener M. Classification and regression by randomForest. R News 2002;2:18-22.
44. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12:77.
45. Hunter JD. Matplotlib: a 2D graphics environment. Computing in Science and Engineering 2007;9:90-5.
46. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer; 2016.
47. Konopka T. umap: uniform manifold approximation and projection. R package. Github, v0.2.10.0, February 1, 2023. Available at: https://github.com/tkonopka/umap.
48. Plotly Technologies. Collaborative data science. Montreal: Plotly Technologies; 2015. Available at: https://plot.ly.
49. Martinson AK, Chin AT, Butte MJ, Rider NL. Artificial intelligence and machine learning for inborn errors of immunity: current state and future promise. J Allergy Clin Immunol Pract 2024;12:2695-704.
50. Mendez Barrera JA, Rocha Guzman S, Hierro Cascajares E, Garabedian EK, Fuleihan RL, Sullivan KE, et al. Who's your data? Primary immune deficiency differential diagnosis prediction via machine learning and data mining of the USIDNET registry. Clin Immunol 2023;255:109759.
51. FitzPatrick AM, Chin AT, Nirenberg S, Cunningham-Rundles C, Sacco K, Perlmutter J, et al. Piloting an automated query and scoring system to facilitate APDS patient identification from health systems. Front Immunol 2024;15:1508780.
52. Bastarache L, Hughey JJ, Hebbring S, Marlo J, Zhao W, Ho WT, et al. Phenotype risk scores identify patients with unrecognized Mendelian disease patterns. Science 2018;359:1233-9.
53. Jimenez-Mesa C, Ramirez J, Yi Z, Yan C, Chan R, Murray GK, et al. Machine learning in small sample neuroimaging studies: novel measures for schizophrenia analysis. Hum Brain Mapp 2024;45:e26555.
54. Villaescusa-Navarro F, Anglés-Alcázar D, Genel S, Spergel DN, Somerville RS, Dave R, et al. The CAMELS project: cosmology and astrophysics with machine learning simulations. Astrophysical Journal 2021;915:71. https://doi.org/10.3847/1538-4357/abf7ba.
55. Barash M, McNevin D, Fedorenko V, Giverts P. Machine learning applications in forensic DNA profiling: a critical review. Forensic Sci Int Genet 2024;69:102994.
56. Achar SK, Keith JA. Small data machine learning approaches in molecular and materials science. Chemical Reviews 2024;124:13571-3.
57. Fox TA, Massey V, Lever C, Pearce R, Laurence A, Grace S, et al. Pre-transplant immune dysregulation predicts for poor outcome following allogeneic haematopoietic stem cell transplantation in adolescents and adults with inborn errors of immunity (IEI). J Clin Immunol 2025;45:64.